

Association of Language Testers in Europe (ALTE)

Handreichungen für Testautoren

(1995, Überarbeitet Juli 2005)

Deutsche Übersetzung:
Dr. Sibylle Bolton, August 2006

VORWORT

Goethe-Institut und TestDaF-Institut haben sich selbst verpflichtet, die Qualitätsstandards der Association of Language Testers in Europe – ALTE – für die Erstellung, Durchführung und Bewertung von Sprachtests, einzuhalten¹. Beide Institutionen sind bereit, ihre Qualitätssicherungsmaßnahmen durch Gutachter evaluieren zu lassen und legen großen Wert auf die Erprobung und fortwährende Überprüfung von Testaufgaben und – ergebnissen. Sie haben jeweils ein umfangreiches Programm zur Schulung und Fortbildung von Testautoren, Beurteilern und Prüfern. Ziel dieser Maßnahmen zur Qualitätssicherung sind objektive, reliable, valide und faire Tests. Goethe-Institut und TestDaF-Institut setzen sich für die Verbreitung solcher Qualitätsstandards in der deutschsprachigen Fachwelt ein.

Diese *Handreichungen für Testautoren* wurden von der University of Cambridge ESOL (English for Speakers of Other Languages) in englischer Sprache unter dem Titel *Item Writer Guidelines* entwickelt und erstmals 1995 veröffentlicht. Im Jahr 2005 wurde diese Fassung in völlig überarbeiteter Form neu herausgebracht und den ALTE-Partnern zur Verfügung gestellt. Auf der ALTE-Homepage sind sie auch für die Öffentlichkeit zugänglich². Die vier Module der Handreichungen sind als praktischer Leitfaden für die verschiedenen Phasen der Testentwicklung konzipiert, sie werden durch Literaturempfehlungen ergänzt und vertieft. Die Module können im Ganzen, aber auch einzeln für Seminare und Schulungen verwendet werden.

Die deutsche Fassung zitiert die Aufgaben- und Kandidatenbeispiele der englischen Originalausgabe. Sämtliche Erklärungen oder Bewertungen, auch Bewertungskriterien und Bewertungsanleitungen sind jedoch ins Deutsche übersetzt. Diese Handreichungen enthalten darüber hinaus Hinweise und Tipps, die sich auf die Praxis der Prüfungsentwicklung bei ESOL beziehen. Wo es erforderlich erschien, wurden sie durch Anmerkungen in der Übersetzung erläutert.

Die Übersetzung der Handreichungen ins Deutsche übernahm Sibylle Bolton (Goethe-Institut); Redaktion und Produktion lagen bei Gabriele Kecker (TestDaF-Institut). Beiden sei an dieser Stelle für ihre genaue und geduldige Arbeit gedankt.

Dr. Katharina von Ruckteschell-Katte
Goethe-Institut, Zentrale München

Dr. Hans-Joachim Althaus
TestDaF-Institut, Hagen

¹ Vgl.: ALTE Code of Practice (deutsch): http://www.alte.org/quality_assurance/code/code_practice_de.pdf

² ALTE Materials for the Guidance of Test Item Writers (2005), http://www.alte.org/projects/item_writer.php

INHALTSVERZEICHNIS

Allgemeine Einführung in den Kurs

MODUL 1

MODELLE DER SPRACHKOMPETENZ

1. DER PSYCHOMETRISCH-STRUKTURALISTISCHE ANSATZ

- a) Kritik am strukturalistischen Ansatz

2. DIE PRAGMATISCHE WENDE

3. DER BEITRAG DES THRESHOLD LEVEL

4. MODELLE KOMMUNIKATIVER KOMPETENZ

- a) Das Modell von Canale und Swain
- b) Cummins Modell der kommunikativen Kompetenz
- c) Morrow und authentische Kommunikation
 - i) Nachteile von Morrows Ansatz
- d) Bachman und die kommunikative Kompetenz
 - i) Änderungen am Bachman Modell

5. VALIDITÄT

- a) Verfahren zur Ermittlung der Validität
- b) Aspekte der Validität und wie man sie erreichen kann
 - i) Inhaltliche Validität
 - ii) Kriterienbezogene Validität
 - iii) Konstruktvalidität
 - iv) Augenscheinvalidität
- c) Neuere Ansichten zur Testvalidität
 - i) Die Matrix von Messick
- d) Praxisorientierte Rahmenmodelle für die Testvalidierung
 - i) Nachweiszentrierte Testentwicklung: Mislevy
 - ii) Rahmenmodell zur Validierung von Tests (Weir)

6. ÜBUNGEN

ANHANG A – Literaturempfehlungen

ANHANG B – Bibliographie

MODUL 2

TESTENTWICKLUNG

1. CHECKLISTE FÜR DIE BEURTEILUNG VON TESTS

2. TESTBESCHREIBUNG (ENGL. SPECIFICATIONS)

3. PRODUKTIONSPROZESS

- a) Auftragsvergabe
 - i) Hinweise zur Auswahl der Texte
 - ii) Hinweise zur äußeren Form
 - iii) Detaillierte Hinweise zu jeder Aufgabe
- b) Begutachtung und Redaktion
- c) Erprobung
- d) Statistische Analysen der Items
 - i) Schwierigkeit der Items
 - ii) Trennschärfe der Items
 - iii) Angaben zu den Distraktoren
- e) Kalibrierung der Items
 - i) Erstellen einer Schwierigkeitsskala
 - ii) Verankerung
 - iii) Interpretation der Scores
- f) Itembank
 - i) Itembank und Testzusammenstellung
 - ii) Beispiel für eine Itembank-Software
 - iii) Testzusammenstellung ohne Itembank

4. ÜBUNGEN

ANHANG A – Abgelehnte Texte

ANHANG B – Iteman-Beispiele (MicroCAT) für Übung 4

ANHANG C – Literaturempfehlungen

ANHANG D – Ein Beispiel für die Testentwicklung,
Entwicklung eines Tests für die Fertigkeit Sprechen

MODUL 3

ITEMTYPEN

EINIGE ASPEKTE DER ITEMERSTELLUNG

1. TEXTE

- a) Authentizität
 - i) situationsbezogene Authentizität
 - ii) interaktionsbezogene Authentizität
- b) Schwierigkeit der Texte
 - i) Sprachliche Struktur des Textes
 - ii) Kontextuelle Einbettung des Textes
 - iii) Inhalt des Textes
 - iv) Art der Interaktion
- c) Schwierigkeit in Hörtexten
 - i) Interaktion der Sprecher
 - ii) Zeitliche Bezüge und Kontext
 - iii) Sprache

2. ITEMTYPEN

- a) Einige allgemeine Regeln
- b) Multiple-Choice- und andere Auswahlitems
 - i) Einzelne und textbezogene Multiple-Choice-Items
 - ii) Richtig/Falsch-Items
 - iii) Lückentest mit Multiple-Choice-Items
 - iv) Lückentest mit Schüttelkasten
 - v) Textabschnitte zuordnen
 - vi) Zuordnungsaufgaben
 - vii) Mehrfachzuordnung
 - viii) Fehlerkorrektur: Ein Wort zu viel
- c) Items mit subjektivem Beantwortungsspielraum
 - i) Kurzantwortaufgaben
 - ii) Satzergänzung
 - iii) Lückentest mit offenen Lücken
 - iv) Umformungen
 - v) Wortbildung
 - vi) Lückentest zur Wortbildung
 - vii) Aus Stichwörtern einen Text erstellen
 - viii) Korrektur lesen/Fehler korrigieren
 - ix) Informationstransfer

3. NICHT-ITEMBASIERTE AUFGABENTYPEN

- a) Schreiben: Aufgaben zum erweiterten Schreiben
 - i) Aufgaben zum Schreiben mit detailliertem Input
 - ii) Aufgaben zum Schreiben mit geringem Input
 - iii) Beispiele für Aufgaben zum Schreiben auf unterschiedlichen Stufen
- b) Sprechen: Aufgaben zum Sprechen
 - i) Vortrag
 - ii) Vorgabe von Bildern
 - iii) Schriftliche Vorgaben
 - iv) Aufgaben mit einer Informationslücke

4. ANWEISUNGEN

5. LÖSUNGSSCHLÜSSEL, BEWERTUNGSVORGABEN,
BEWERTUNGSANLEITUNG

6. ÜBUNGEN

MODUL 4

ASPEKTE DER BEWERTUNG UND AUSWERTUNG

1. EIN FAIRES ERGEBNIS ERREICHEN

2. RELIABILITÄT

- a) In objektiv ausgewerteten Tests
 - i) Interne Faktoren
 - ii) Externe Faktoren
- b) In subjektiv bewerteten Tests

3. EINIGE ASPEKTE BEI DER BEWERTUNG UND AUSWERTUNG OBJEKTIVER TESTVERFAHREN

- a) Wie können Tests mit objektiven Testverfahren ausgewertet werden?
 - i) Auswertung per Computer
 - ii) Auswertung mit Antwortschablone
- b) Rohwerte in Testergebnisse umrechnen

4. EINIGE ASPEKTE DER SUBJEKTIVEN LEISTUNGSBEWERTUNG

- a) Wie kann man die produktiven Fertigkeiten fair bewerten?
 - i) Die Materialien und das Testformat
 - ii) Die Bewertungsverfahren
- b) Tests zum Schreiben
 - i) Bewertungsverfahren
 - ii) Beurteilertraining
 - iii) Mehrfachkorrektur
- c) Tests zum Sprechen
 - i) Worin unterscheidet sich die Bewertung schriftlicher Leistungen von der Bewertung mündlicher Leistungen?
 - ii) Aus welchen Phasen besteht ein Prüfertraining?
 - iii) Aspekte der Fairness bei der Bewertung mündlicher Leistungen

ANHANG A – Formeln zur Ermittlung der Reliabilität eines Tests

ANHANG B – Beispiel für einen Subtest mit Bewertungsvorgabe

ANHANG C – Beispiel für einen Subtest zum Schreiben mit Bewertungsanleitungen und Musterkorrekturen

ALLGEMEINE EINFÜHRUNG IN DEN KURS

An wen wendet sich dieser Kurs?

Die vorliegenden Materialien dienen der Fortbildung und wenden sich an alle, die mit Tests befasst sind, sei es mit der Entwicklung von Fremdsprachentests, dem Erstellen von Testaufgaben, der Testdurchführung oder der Bekanntgabe der Ergebnisse.

In vielen Fällen, in denen Lehrer* Tests entwickeln müssen, um den Lernfortschritt ihrer Lerner festzustellen, ist es ein und dieselbe Person, die in jeder Phase des Prozesses tätig wird, womöglich ohne zusätzliche personelle Unterstützung. In anderen Situationen, in denen weit verbreitete, staatlich anerkannte und evtl. kommerziell vertriebene Feststellungsprüfungen oder Kursabschlussprüfungen zum Einsatz kommen, handelt es sich möglicherweise nur um die Mitarbeit in einer der Phasen, z. B. als Testautor oder als Prüfer. Diese Handreichungen beziehen sich auf alle diese Situationen.

Wie ist der Kurs aufgebaut?

Um die verschiedenen Phasen und Inhalte widerzuspiegeln, die den Prozess der Testentwicklung und -durchführung ausmachen, und um auf die unterschiedlichen Interessen der daran Beteiligten einzugehen, sind die Materialien in vier Module unterteilt:

In Modul 1 geht es um die Frage, was ein Test prüfen soll, d. h., es werden verschiedene Modelle der Sprachkompetenz vorgestellt.

In Modul 2 geht es um den Prozess der Testentwicklung.

In Modul 3 werden unterschiedliche Aufgabentypen beschrieben.

In Modul 4 geht es um die Bewertung und Auswertung der Prüfungsergebnisse.

Jedes Modul kann unabhängig von den anderen bearbeitet werden. Die Aufteilung in Module bedeutet jedoch nicht, dass ein Mitarbeiter, der z. B. als Testautor nur mit dem Schreiben von Items befasst ist, sein Wissen über den Prozess der Testentwicklung und -durchführung nur auf diesen einen Bereich beschränken sollte. Für Testautoren ist es von Bedeutung, die verschiedenen Aspekte der Sprachkompetenz sowie das dem Test zugrunde liegende Konzept zu kennen, da beides die Auswahl der Aufgabentypen bestimmt, die sie entwickeln sollen. Außerdem ist es für Testautoren wichtig, über alle Aspekte und Fragen der Auswertung und Bewertung von Prüfungsergebnissen Bescheid zu wissen.

Die vier Module behandeln verschiedene Phasen und unterschiedliche Aspekte der Testentwicklung und -durchführung, sind aber jeweils als Teil eines Ganzen zu sehen.

Diese Handreichungen enthalten Informationen, Überlegungen, Anregungen und Beispiele, die sich auf Fremdsprachentests beziehen und die möglichst vielen an der Testentwicklung und -durchführung Beteiligten von Nutzen sein sollen. Es wird versucht, aus den Erfahrungen, die bei der Arbeit in einzelnen Testinstitutionen gewonnen wurden, allgemeingültige Prinzipien abzuleiten, die für Fremdsprachentests in den verschiedensten Testsituationen gelten. Diese Überlegungen und Anregungen reflektieren nicht notwendigerweise die Praxis einer bestimmten Prüfungsinstitution, es sind jedoch Prinzipien, die sich in der Praxis als fundiert erwiesen haben.

Wie kann der Kurs eingesetzt werden?

Der Kurs kann entweder im Selbststudium bearbeitet oder als Unterrichtsmaterial in Kleingruppen oder Seminaren eingesetzt werden. Übungsaufgaben und Lösungsvorschläge sind an geeigneten Stellen integriert.

* Anmerkung (Übersetzerin): Aus Gründen der sprachlichen Vereinfachung werden in den Modulen Ausdrücke wie „Beurteiler“, „Prüfungsteilnehmer“, „Testentwickler“ usw. im generischen Sinne verwendet.

Terminologie

Verschiedene Testinstitutionen oder Experten verwenden unterschiedliche Terminologien, wenn sie von verschiedenen Teilen eines Tests oder den verschiedenen Aspekten des Testens sprechen. Um für den Kurs eine einheitliche Verwendung sicherzustellen, werden einige wichtige und häufig gebrauchte Begriffe wie folgt definiert:

SUBTEST	Ein Test besteht üblicherweise aus einzelnen Subtests zu den verschiedenen Fertigkeiten, die nach diesen Fertigkeiten benannt werden. Man kann sie auch als Prüfungsteil bezeichnen.
INPUT	Das Material, das den Kandidaten vorgegeben wird, damit sie in angemessener Weise sprachlich darauf reagieren. Man kann es auch als Testvorgabe bezeichnen.
ITEM	Jedes Einzelelement eines Tests, das getrennt bewertet wird. Das kann z. B. eine Lücke in einem schriftlichen Text sein oder ein Multiple-Choice-Item mit drei oder vier Antwortoptionen.
ANTWORT	Die schriftlichen oder mündlichen Reaktionen der Kandidaten auf die Testaufgaben.
ANWEISUNG	Instruktionen, die den Kandidaten gegeben werden, um ihre Antwort auf eine bestimmte Testaufgabe zu steuern.
(TEST-)TEIL	Die Teile, in die ein Subtest unterteilt werden kann. Jeder Teil kann einen unterschiedlichen Aspekt der jeweiligen Fertigkeit oder des Sprachsystems testen.
AUFGABE	Eine Aufgabe wird definiert durch die Kombination von Anweisung, Input und Antwort. So kann z. B. ein schriftlicher Text mit mehreren Multiple-Choice-Items, die alle mit Bezug auf eine Anweisung gelöst werden können, als Aufgabe bezeichnet werden. Wenn sich weitere Items ebenfalls auf diesen Text beziehen, aber eine andere Anweisung zur Lösung vorgeben, wird diese Kombination aus Text, Anweisung und Items als separate Aufgabe angesehen. Derselbe Text kann also als Input für eine oder mehrere Aufgaben dienen. Der Begriff „Frage“, der manchmal statt „Aufgabe“ verwendet wird, unabhängig davon, ob die Aufgabe als Frage formuliert ist, sollte vermieden werden, da er nicht eindeutig ist.
TEXT	Jede zusammenhängende mündliche oder schriftliche Äußerung, die als Input in einem Test verwendet wird oder die als Output vom Kandidaten produziert wird.